BAB I
PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, pembelajaran online semakin populer dan menjadi pilihan utama banyak orang. Pembelajaran online sendiri berarti metode pendidikan yang menggunakan teknologi yang memungkinkan siswa berinteraksi secara daring melalui fasilitas jaringan internet (Yanti Nasution et al., 2022). Istilah online learning atau pembelajaran online banyak disinonimkan dengan istilah lainnya seperti e-learning, internet learning, web-based learning, tele-learning, distributed learning dan lain sebagainya (Belawati, 2019). Dalam konteks penelitian ini fokus dari media pembelajaran online yang digunakan adalah video pembelajaran

Dalam konteks pembelajaran online melalui media video yang dapat diakses secara online dapat menggunakan platform seperti YouTube. Penggunaan aplikasi seperti *speech-to-text* dapat menjadi teknologi yang akan sangat berguna dalam mendukung transkripsi otomatis dari konten video. Dengan adanya aplikasi ini, proses transkripsi dapat dilakukan secara otomatis, untuk mengubah ucapan yang terdengar dalam video menjadi teks yang tertulis. Hal ini memiliki banyak manfaat, terutama bagi peserta pembelajaran online, yang dapat memanfaatkan teks hasil transkripsi untuk membantu pemahaman materi yang ada dalam konten secara lebih baik.

Speech recognition sendiri berkaitan erat dengan sistem konversi *speech to text* yang memungkinkan pembuatan transkripsi otomatis dari ucapan yang ada dalam video yang memiliki output audio. Dalam konteks ini, akurasi sistem *speech recognition* adalah faktor kunci yang bisa memengaruhi pemahaman dan efektivitas pembelajaran online. Meskipun telah ada kemajuan dalam teknologi *speech recognition*, masih terdapat pertanyaan tentang tingkat akurasi dan sejauh mana teknologi ini dapat digunakan dalam berbagai seperti pembelajaran online melalui video di YouTube.

Dari peneltian terdahulu terkait dengan teknologi *speech recognition* yaitu, Implementasi Aplikasi *Speech to Text* untuk Memudahkan Wartawan Mencatat Wawancara dengan Python. Dalam penelitian, *speech recognition* digunakan untuk proses mengubah suara menjadi teks dalam pengembangan aplikasi. Untuk menyelesaikan masalah tersebut, dibuat aplikasi yang menggunakan bahasa pemrograman Python untuk mengkonversi suara ke teks atau tulisan. Pada

tahap uji coba yang menggunakan 6 sample rekaman suara dan menunjukkan tingkat keberhasilan 94,75% dalam mengkonversi suara ke tulisan.(Buana, 2020).

Pada penelitian lain yang berfokus pada pengembangan aplikasi sejenis *Speech to Text* pada kasus perancangan Aplikasi Pembelajaran Siswa Berbasis Web Menggunakan *Speech To Text* Pada Sdn 2 Pabuaran. Hasilnya sebuah aplikasi yang dapat membantu siswa maupun guru dalam proses pembelajaran seperti mendapatkan materi atau memberikan materi pembelajaran. (Hidayat & Mulyoto, 2022). *Speech recognition* digunakan oleh aplikasi *speech to text* berbasis web yang didukung Javascript. *Web Speech API* memungkinkan pengembang web memasukkan *speech recognition* dan sintesis ke dalam aplikasi web mereka. API ini memungkinkan mereka untuk merekam input audio, mengidentifikasi ucapan, dan konversi ucapan ke teks. Sistem ini tersedia dalam bentuk JavaScript dengan file HTML dan CSS yang dinamis (- & -, 2023).

Salah satu algoritma yang dapat digunakan untuk mengetahui kecocokan suara dalam penelitian *speech recognition* adalah *Dynamic Time Warping* (DTW). DTW adalah metode untuk mengukur kemiripan suatu pola suara dengan zona waktu yang berbeda. Semakin kecil jarak yang dihasilkan, semakin semakin mirip antara kedua pola suara tersebut. Kedua pola suara tersebut mirip, maka kedua suara tersebut dikatakan sama. Volume pengucapan, durasi pengucapan, dan kebisingan dari suara di sekitar tempat perekaman berlangsung mempengaruhi jarak yang dihasilkan. Durasi dan volume mempengaruhi mempengaruhi jarak yang dihasilkan. Semakin dekat jarak pengucapan data uji dengan data latih maka jarak yang dihasilkan akan semakin kecil. Semakin tinggi volume data uji maka semakin besar jarak yang dihasilkan (Permanasari et al., 2019).

Penggunaan transkripsi audio yang dilakukan secara manual dapat memiliki tingkat akurasi yang lebih tinggi tetapi hal tersebut akan memakan waktu dan sumber daya. Oleh karena itu, solusi teknologi yang lebih efisien diperlukan untuk mengotomatisasi proses ini. Dalam konteks pembelajaran online, penerapan teknologi *speech to text* atau *speech recognition* dapat membantu menyederhanakan proses transkripsi serta meningkatkan efisiensi waktu dan tenaga. Dengan teknologi ini, materi pembelajaran dapat secara otomatis dan instan diubah menjadi teks pada saat penyampaiannya, sehingga peserta pembeajaran dapat lebih memahami materi pembelajaran.

Dalam penelitian terkait pengukuran akurasi, Penerapan Metode *Mel Frequency Cepstral Coefficients* Pada Sistem Pengenalan Suara Berbasis Desktop. Metode yang digunakan dalam penelitian ini adalah Metode *Mel Frequency Cepstral Coefficients*. Pada hasil pengujian,

menunjukkan bahwa pada kondisi pengujian ideal, sistem mempunyai tingkat keberhasilan sebesar 90% dan tingkat kegagalan sistem sebesar 10% dengan tingkat kesalahan atas level 5 sebesar 0%. ketika diuji dalam kondisi tidak ideal, sistem mempunyai tingkat keberhasilan sebesar 76,6667% dan tingkat kegagalan sistem sebesar 23,333%, dengan lima tingkat kegagalan teratas adalah 0% (Ajinurseto & Islamuddin, 2023).

Oleh karena itu, evaluasi keakuratan sistem *speech recognition* dalam konteks pembelajaran online sangat penting. Penelitian sebelumnya telah memberikan wawasan tentang penggunaan teknologi ini dalam konteks yang berbeda, sehingga penelitian lebih lanjut diperlukan dengan metode lainnya. Penelitian ini memiliki tujuan untuk memberikan pemahaman yang mendalam tentang seberapa baik teknologi *speech recognition* membantu pembelajaran online melalui video dengan mengukur tingkat akurasi dan membandingkannya dengan transkripsi manual.

Diperlukan juga sebuah metode yang dapat digunakan untuk mengevaluasi atau menganalisa pengukuran akurasi antara teknologi konversi *speech to text* dengan transkripsi manual yaitu menggunakan metode perhitungan *Word Error Rate* (WER). Kata yang disubstitusi, dihapus, dan ditambahkan adalah tiga jenis kesalahan utama yang digunakan sebagai parameter dalam WER. *Word Error Rate* (WER) digunakan untuk menentukan persentase keberhasilan konversi ucapan ke teks (Dwijayanti et al., 2021). Hal ini dilakukan dengan menghitung jumlah penambahan, penghapusan, dan pensubstitusian dari kata yang digunakan untuk mengonversi ucapan menjadi teks dan dibandingkan dengan total kata dari transkripsi manual.

Penelitian lain terkait penggunaan WER sebagai tolak ukur kinerja system *speech recognition* yang berjudul *Automatic Speech Recognition* Bahasa Indonesia menggunakan *Unidirectional Gated Recurrent Unit*. Penelitian ini menganalisis kemampuan Unidirectional GRU dalam memetakan sinyal akustik kedalam teks. Tolak ukur kinerja dari model diukur menggunakan satuan WER. Data yang digunakan adalah data audio dan transkrip terjemahan bahasa Indonesia dari Al-Quran. Hasilnya jaringan kemudian dioptimasi menggunakan tools Adam optimizer dan menghasilkan Tingkat WER 90.611 % dari model terbaik yang diuji (Firmansyah & Bachtiar, 2021).

Maka dengan latar belakang masalah diatas dibuatlah penelitian dengan judul "Implementasi Teknologi Konversi *Speech-To-Text* Dalam Transkripsi Pembelajaran Online Dengan Algoritma *Dynamic Time Warping*". Dengan tujuan agar mendapat pemahaman yang lebih

baik tentang tingkat akurasi *sistem Speech-to-Text* atau *speech recognition*. Selain itu, dengan penggunaan *speech recognition*, peserta pembelajaran dapat memiliki pengalaman belajar yang lebih baik dan lebih efisien, mengurangi kebutuhan untuk transkripsi manual dan meningkatkan keterlibatan dalam pembelajaran melalui media konten video.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang ada dapat dibuat identifikasi masalah sebagai berikut :

1. Pada implementasi teknologi *speech to text* dalam lingkungan pembelajaran online, seberapa akurat teknologi ini dalam mengidentifikasi ucapan dalam video pembelajaran yang ada di platform youtube ?
2. Sejauh mana akurasi teknologi transkripsi *speech to text* dapat dibandingkan dengan transkripsi manual dalam pembelajaran video?

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang ada dapat dibuat beberapa tujuan dari penelitian sebagai berikut :

1 Untuk mengukur keakuratan teknologi *speech to text* untuk mengenali dan mentranskripsi ucapan dalam video pembelajaran yang ada di platform YouTube.

2 Penelitian ini juga akan membuat perbandingan hasil transkripsi otomatis dengan transkripsi manual. Untuk mencari tahu sejauh mana teknologi dapat menggantikan pekerjaan manusia dalam transkripsi dengan membandingkan hasil transkripsi otomatis dan manual.

1.4 Batasan Masalah

Adapun Batasan masalah yang ada pada penelitian ini sebagai berikut :

1. Batasan penelitian ini adalah hanya mencakup bahasa atau aksen tertentu yaitu Bahasa Indonesia.
2. Durasi video yang digunakan dalam penelitian ini akan dibatasi agar fokus tetap pada akurasi *speech recognition* dalam konteks situasi pembelajaran yang lebih pendek.
3. Sistem *Speech Recognition* yang untuk konversi *speech-to-text* pada penelitian akan menggunakan service dari *Web Speech API*.

1.5 Kontribusi Penelitian

Hasil penelitian ini diharapkan dapat membantu dalam kemajuan teknologi *speech to text* (*STT*) yang lebih baik, khususnya dalam hal pembelajaran online. Dengan mengetahui seberapa baik teknologi STT saat ini mampu mengenali dan mentranskripsi ucapan dalam video

pembelajaran, pengembang teknologi dapat menemukan masalah yang mengganggu kinerja sistem.